Prompting Large Language Model for Machine Translation: A Case Study

Abstract

Research on prompting has shown excellent performance with little or even no supervised training across many tasks. However, prompting for machine translation is still under-explored in the literature. We fill this gap by offering a systematic study on prompting strategies for translation, examining various factors for prompt template and demonstration example selection. We further explore the use of monolingual data and the feasibility of cross-lingual, cross-domain, and sentence-to-document transfer learning in prompting. Extensive experiments with GLM-130B (Zeng et al., 2022) as the testbed show that 1) the number and the quality of prompt examples matter, where using suboptimal examples degenerates translation; 2) several features of prompt examples, such as semantic similarity, show significant Spearman correlation with their prompting performance; yet, none of the correlations are strong enough; 3) using pseudo parallel prompt examples constructed from monolingual data via zero-shot prompting could improve translation; and 4) improved performance is achievable by transferring knowledge from prompt examples selected in other settings. We finally provide an analysis on the model outputs and discuss several problems that prompting still suffers from.